ORIGINAL ARTICLE

Spatio-temporal deep learning models for tip force estimation during needle insertion

Nils Gessert[1] · Torben Priegnitz[1] · Thore Saathoff[1] · Sven-Thomas Antoni[1] · David Meyer[2] ·
Moritz Franz Hamann[2] · Klaus-Peter Jünemann[2] · Christoph Otte[1] · Alexander Schlaefer[1]

Abstract

Purpose Precise placement of needles is a challenge in a number of clinical applications such as brachytherapy or biopsy. Forces acting at the needle cause tissue deformation and needle deflection which in turn may lead to misplacement or injury. Hence, a number of approaches to estimate the forces at the needle have been proposed. Yet, integrating sensors into the needle tip is challenging and a careful calibration is required to obtain good force estimates.

Methods We describe a fiber-optic needle tip force sensor design using a single OCT fiber for measurement. The fiber images the deformation of an epoxy layer placed below the needle tip which results in a stream of 1D depth profiles. We study different deep learning approaches to facilitate calibration between this spatio-temporal image data and the related forces. In particular, we propose a novel convGRU-CNN architecture for simultaneous spatial and temporal data processing.

Results The needle can be adapted to different operating ranges by changing the stiffness of the epoxy layer. Likewise, calibration can be adapted by training the deep learning models. Our novel convGRU-CNN architecture results in the lowest mean absolute error of 1.59 ± 1.3 mN and a cross-correlation coefficient of 0.9997 and clearly outperforms the other methods. Ex vivo experiments in human prostate tissue demonstrate the needle's application.

Conclusions Our OCT-based fiber-optic sensor presents a viable alternative for needle tip force estimation. The results indicate that the rich spatio-temporal information included in the stream of images showing the deformation throughout the epoxy layer can be effectively used by deep learning models. Particularly, we demonstrate that the convGRU-CNN architecture performs favorably, making it a promising approach for other spatio-temporal learning problems.

Keywords Force estimation · Optical coherence tomography · Convolutional GRU · Convolution neural network · Needle placement

Introduction

For minimally invasive procedures such as biopsy, neurosurgery or brachytherapy, needle insertion is often utilized to minimize tissue damage [1]. To facilitate accurate needle placement, needle steering, image guidance and force esti-

✉ Nils Gessert
nils.gessert@tuhh.de

[1] Institute of Medical Technology, Hamburg University of Technology, Hamburg, Germany

[2] Department of Urology, University Hospital Schleswig-Holstein, Kiel, Germany

mation [2] can be used. Accurate measurement of the forces affecting the needle tip is of particular interest, e.g., to keep track of the needle–tissue interaction and to detect potential tissue ruptures, or to generate haptic and visual feedback [3]. Therefore, various force-sensing solutions for needles have been proposed. A simple approach is to place a force sensor externally at the needle shaft which would allow for the use of conventional force–torque sensors. However, during insertion, large frictional forces act on the needle shaft which mask the actual tip forces. Therefore, forces acting on the needle shaft either need to be decoupled from the force sensor or the force sensor needs to be placed at the needle tip [4]. This complicates building force sensors as they are usually constrained to a few millimeters in width [5] which is particularly difficult for mechatronic force sensors [4,6]. For these reasons, fiber-optic force sensors have been developed

which are often smaller, biocompatible and MRI-compatible [7]. In particular, sensor concepts using Fabry-Pérot interferometry [7,8] or Fiber Bragg Gratings [9–11] have been proposed. These methods have shown promising calibration results; however, manufacturing and signal processing can be difficult when different temperature ranges and lateral forces need to be taken into account. Another optical method uses the imaging modality optical coherence tomography (OCT) to estimate strain and deformation from images of deformed material such as silicone [12,13]. Also, direct force estimation from volumetric OCT data has been presented [14,15]. Other approaches have integrated single OCT fibers that produce 1D images into needle probes [16]. This concept has been used to classify malignant and benign tissue [17].

In this work, we present an OCT needle concept for force estimation at the needle tip. A single OCT fiber is embedded into a ferrule with an epoxy layer applied on top of it. A sharp metal tip is mounted on top of the epoxy layer to facilitate tissue insertion. Axial forces acting on the needle tip lead to a deformation of the epoxy layer which is imaged by the OCT fiber. Thus, forces can be inferred from the OCT signal. In general, this needle design is easy to manufacture and flexible as no precise fiber placement is required; the needle tip's shape can be changed and the epoxy layer's thickness and composition can be varied. Thus, softer epoxy resin could be used for application scenarios which require a high sensitivity such as microsurgery and stiffer epoxy resin could be used for large forces which occur, e.g., during biopsy [7]. However, this approach comes with challenges for calibration and force estimation. In particular, a robust, nonlinear model is required which maps the deformations observed in the OCT images to forces. The signal can be understood as 2D spatio-temporal data with a spatial and a temporal dimension. Using the current observation t_i and previous ones has been shown to be effective for vision-based force estimation with RGBD cameras as the current force estimate is likely reflected in prior deformation [18]. In contrast to previous approaches [18], we directly learn relevant features from the data using deep learning models. This eliminates the necessity to engineer new features for other materials with different light scattering properties.

For deep learning-based spatio-temporal data processing, various method have been proposed in the natural image domain, e.g., for action recognition [19,20] or video classification [21,22]. Spatial and temporal convolutions have been employed [23,24] and models using convolutional neural networks (CNNs) with a subsequent recurrent part have been proposed [21,25]. We adapt these approaches for our new 2D spatio-temporal learning problem where a series of 1D OCT A-scans needs to be processed. In addition, we propose a novel convolutional gated recurrent unit-convolutional neural network (convGRU-CNN) architecture. The key idea and difference to other methods is to first learn temporal relations using recurrent layers while keeping the spatial data structure intact by using convolutions for the recurrent gating mechanism [26]. Then, a 1D CNN architecture processes the resulting spatial representation. We provide an in-depth analysis of this concept and compare it to previous approaches for deep learning-based spatio-temporal data processing. In addition, we provide qualitative results for tissue insertion experiments, showing the feasibility of the approach. This work extends preliminary results we presented at MICCAI 2018 [27]. We substantially revised and extended the original paper with an extended review of the relevant literature, a more detailed explanation of our novel convGRU-CNN model and additional experiments for the model. In particular, we rerun all quantitative experiments for more consistent results, we perform additional experiments to analyze the temporal dimension and properties of the convGRU-CNN and we improve the model with recurrent batch normalization [28] and recurrent dropout [29]. To highlight the proposed model's advantages, we consider additional spatio-temporal deep learning models and a conventional model. We provide a more detailed comparison of the spatio-temporal models' errors and their significance. Furthermore, we provide inference times of our models to demonstrate real-time capability.

Summarized, the key contributions of this work are threefold. (1) We propose a new design for OCT-based needle tip force estimation that is flexible and easy to manufacture. (2) We present a novel convGRU-CNN architecture for spatio-temporal data processing which we use for calibration of our force-sensing mechanism. (3) We show the feasibility of our approach with an insertion experiment into human ex vivo tissue.

Materials and methods

Problem definition

Our force-sensing needle design uses OCT which produces series of 1D images (A-scans) that need to be mapped to forces. Thus, we consider a 2D spatio-temporal learning problem with a set of t_s consecutive, cropped 1D A-scans $(A_{t_i}, A_{t_{i-1}}, \ldots, A_{t_{i-t_s}})$ with $A_{t_i} \in \mathbb{R}^{d_c}$ where d_c denotes the crop size. The resulting matrix $M_{t_i} \in \mathbb{R}^{t_s \times d_c}$ is used to estimate a force $F_{t_i} \in \mathbb{R}$.

Deep learning architectures

We consider different deep learning models to map M_{t_i} to F_{t_i}. First, we introduce our novel convGRU-CNN architecture. Then, we consider model variants that use alternative ways of data processing.

convGRU-CNN combines spatial and temporal data processing in a new way. First, a convolutional GRU (convGRU)
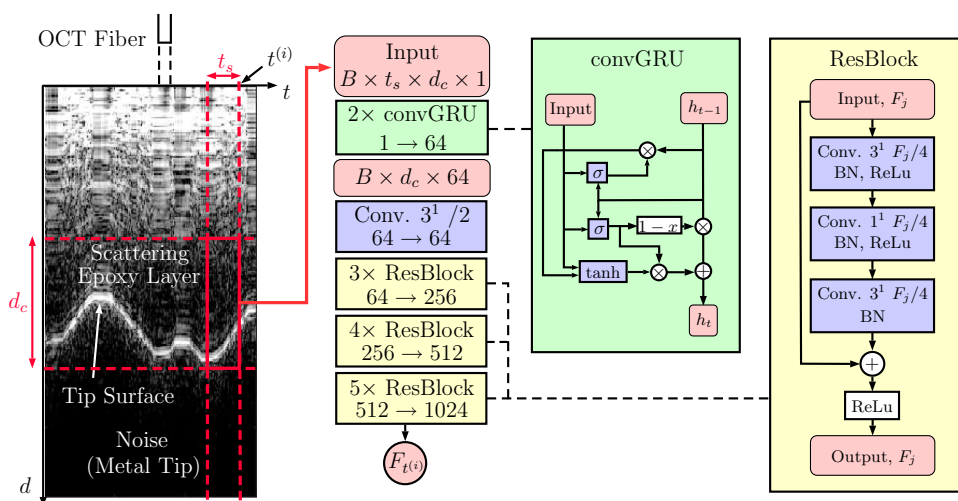
Fig. 1 The convGRU-CNN model we employ. The metal tip's flat surface at the epoxy layer cannot be penetrated by infrared light which is why that signal part is considered noise. σ and tanh denote a convolutional gate with sigmoid and hyperbolic tangent activation function, respectively. The subsequent CNN is a ResNet-like network. The first block in a series of ResBlocks uses a stride of 2 for the convolutions with kernel 3^1 and increases the number of feature maps. Subsequent blocks have a stride of 1 and keep the same feature map size. The change in the number of feature maps F_j is denoted in each group of ResBlocks

takes care of temporal processing. The convGRU outputs a 1D spatial feature representation which is then processed by a ResNet-inspired [30] 1D CNN. The model is shown in Fig. 1. The convGRU is a combination of convLSTM and gated recurrent units [31]. We replace the matrix multiplications in the GRU with convolutions such that the output of the convGRU unit is computed as

$$z_t = \sigma(K_z * h_{t-1} + L_z * \text{RBN}(x_t))$$
$$r_t = \sigma(K_r * h_{t-1} + L_r * \text{RBN}(x_t))$$
$$c_t = \tanh(K_c * (r_t h_{t-1}) + L_c * \text{RBN}(x_t))$$
$$h_t = z_t c_t + (1 - z_t) h_{t-1}$$

(1)

where h is the hidden state, x is the input, K and L are filters, $*$ denotes a convolution, σ denotes the sigmoid activation function and RBN(.) denotes recurrent batch normalization [28]. Furthermore, we employ recurrent dropout for additional regularization [29] at the cell input with probability $p_{di} = 0.1$ and at the cell output with probability $p_{do} = 0.2$. Recurrent batch normalization and dropout are extensions to the original model presented in [27] and the new model is named **convGRU-CNN+**. We add these augmentations to all recurrent models.

1D CNN processes A-scans A_{t_i} individually without considering a history of data which resembles a single-shot learning approach. The CNN architecture is the same ResNet-based model that is depicted in Fig. 1.

GRU processes a set a of A-scans, without taking spatial structure into account as it consists of three GRU layers with standard matrix multiplications being performed inside the gates.

CNN-GRU also uses a set a of A-scans and follows the classic approach of first performing spatial processing and feature extraction with a CNN and then temporal processing with a recurrent model. The CNN part is the ResNet-based model as shown in Fig. 1 and the recurrent part is a two-layer GRU.

2DCNN is fed with a set a of A-scans and performs data processing with convolutions over both the spatial and temporal dimensions. This architecture also follows the ResNet-like CNN part shown in Fig. 1. The kernels are of size 3×3, and strides are used to simultaneously reduce the spatial and the temporal dimensions.

GRU-CNN is a variant of convGRU-CNN with normal GRU cells. Here, the A-scans are directly treated as feature vectors. This architecture is used to demonstrate the necessity of using convolutional GRUs when performing temporal processing first.

CNN-convGRU is a variation of CNN-GRU with convGRU cells. Before the global average pooling operation, the convGRU cells perform temporal processing while keeping the spatial structure that resulted from CNN processing. Afterward, global average pooling is applied and the resulting feature vector is fed into the output layer. This architecture serves as a comparison to convGRU-CNN in terms of the position of the convGRU units in the network.

MIP-GPM is a simple reference model using classic feature extraction with a Gaussian process regression model [32]. We extract the needle tip's high-intensity surface using 1D maximum intensity projection (MIP) on the median-filtered A-scans. The normalized pixel index of the MIP

Fig. 2 Schematic drawing of the needle and the calibration setup. Not to scale. The needle contains an OCT fiber that images a deformable epoxy layer below the needle tip. Forces are measured by the force sensor at the base. The setup is moved with a linear stage
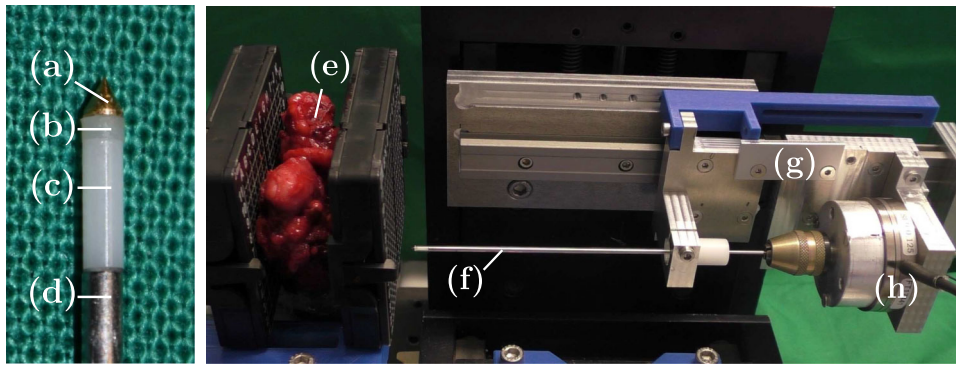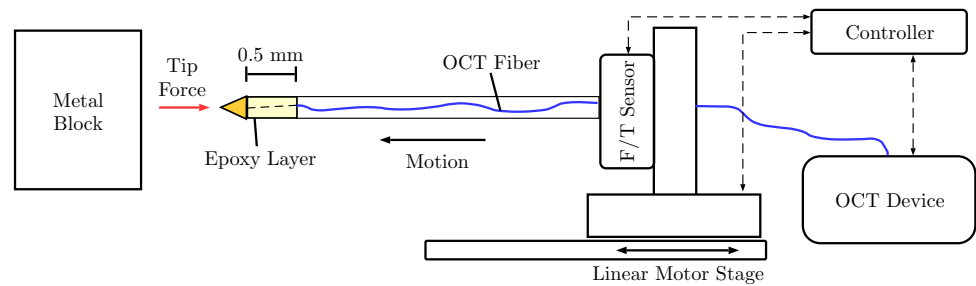

Fig. 3 Needle design (left) and photograph of the experimental setup for the prostate insertion experiment (right). The brass tip (a) is attached to the epoxy layer (b) which is glued to the ferrule with the embedded OCT fiber (c). The ferrule is attached to the needle base (d) with a diam-eter of 1.25 mm. For the prostate (e) insertion experiment, the needle is decoupled with a shielding glass tube (f). The linear stage (g) moves the needle and the force sensor (h) acquires reference data

represents a simple feature that captures deformation. This scalar feature serves as a comparison to the deep learning models.

All networks are trained end-to-end. We use the Adam algorithm for optimization with a batch size of $B = 100$. Our implementation uses Tensorflow [33]. The initial learning rate is $l_r = 10^{-4}$. We halve the learning rate every 30 epochs and stop training after 300 epochs.

Needle design and experimental setup

Our proposed needle tip force-sensing mechanism and calibration setup are shown in Fig. 2. The needle's base is a ferrule with a diameter of 1.25 mm which holds the OCT fiber. On top, we apply an epoxy resin layer with a height of 0.5 mm using Norland Optical Adhesive (NOA) 63. On top of the layer, a cone-shaped brass tip is attached. The epoxy resin layer's stiffness is varied by mixing the resin with different concentrations of NOA 1625. The needle's OCT fiber is attached to a frequency-domain OCT device (Thorlabs Telesto I). A force sensor (ATI Nano43) for ground-truth annotation is mounted between the needle and a linear stage that moves the needle along its axial direction. For calibration, the tip is deformed with random magnitude and velocity to create a large dataset with extensive force variations being covered. Next, we validated the needle in tissue insertion

experiments, see Fig. 3. Obtaining ground-truth tip forces is challenging for this case as the force sensor at the base measures both axial tip forces and friction forces acting on the shaft. Therefore, we use a shielding tube which is decoupled from the needle and the force sensor. This allows for measurement of axial tip forces for comparison to our needle tip sensing mechanism. Note that the shielding tube is a workaround for validation experiments but not for practical application as the stiff tube would increase trauma. We perform insertion experiments into a freshly resected human prostate. In the supplementary material, video, ultrasound, force and OCT signal recordings are provided.

Data acquisition and datasets

The OCT device we use is a frequency-domain-OCT which uses interferometry with near-infrared light to acquire 1D depth profiles (A-scans) with a rate of 5500 Hz. The light's wavelength of 1325 nm allows for imaging of the inner structure of scattering materials with up to 1 mm depth. The force sensor for ground-truth annotation acquires data at 500 Hz. Therefore, the OCT and force sensor data streams need to be synchronized and matched. We use the streams' timestamps for synchronization and nearest neighbor interpolation to assign an A-scan to each force measurement. To construct

Table 1 Comparison of needles with different epoxy layer stiffnesses. The convGRU-CNN+ model was used for this experiment

	MAE	rMAE	CC	Max
Needle 1	1.59 ± 1.3	0.0199 ± 0.0172	0.9997	379
Needle 2	7.12 ± 5.8	0.0266 ± 0.0226	0.9995	974
Needle 3	22.97 ± 20.6	0.0345 ± 0.0301	0.9991	3202

a sequence, we add t_s previous A-scans to each A-scan with an assigned force value.

We acquire calibration datasets for three needles with different stiffness of the epoxy layer. The datasets each contain approximately 90000 sequences of A-scans, each labeled with a scalar, axial force. We use 80 % of the data for training and validation and 20 % for testing. There is no overlap between the sequences from the different sets. We tune hyperparameters based on validation performance. In terms of metrics, we report the mean absolute error (MAE) in mN with standard deviation, the relative MAE (rMAE) with standard deviation and correlation coefficient (CC) between predictions and targets. To ensure consistency, we repeat all experiments five times and provide the mean values over all runs. We test for significant difference in the median of the models' absolute errors with the Wilcoxon signed-rank test ($\alpha = 5$ % significance level). Furthermore, we provide the inference time (IT) in ms of each model for a single forward pass, averaged over 100 repetitions.

Results

First, we report the results for the three different needles with different stiffness of the epoxy layer. Stiffness increase from needle one to three. The results with the corresponding maximum force magnitudes are shown in Table 1. With increasing stiffness, the MAE increases, as the overall covered force range increases. Between needle 1 and 2, the rMAE increases by a factor of 1.33. Between needle 2 and 3, the rMAE increases by a factor of 1.29. The CC remains similar among the needles.

Next, we compare our proposed convGRU-CNN model to other spatio-temporal deep learning methods. The results are shown in Table 2. Overall, the models that perform spatio-temporal processing (convGRU-CNN+, convGRU-CNN, CNN-GRU, CNN-convGRU, 2DCNN) clearly outperform 1DCNN and GRU. Overall, convGRU-CNN+ performs best. Boxplots in Fig. 4 show a more detailed analysis of the spatio-temporal deep learning models. The null hypothesis of an equal median for the absolute errors of convGRU-CNN+ can be rejected with p-values of 6.15×10^{-10}, 1.32×10^{-11}, 7.72×10^{-12} and 1.16×10^{-12} compared to convGRU-CNN, CNN-GRU, CNN-convGRU and 2DCNN, respectively.

In terms of inference time, the spatio-temporal deep learning models can provide predictions with approximately 100 Hz. The fastest spatio-temporal deep learning model is 2DCNN with an IT of 8.6 ms and the overall fastest model is GRU with an IT of 2.5 ms. Note that these values are highly hardware (NVIDIA GTX 1080 Ti) and software (Tensorflow) dependent.

The previous results showed a clear performance increase for joint spatio-temporal processing. Therefore, we perform experiments to analyze the effect of the temporal dimension. In Fig. 5, we show results for different t_s and the associated training durations with our convGRU-CNN+ and CNN-GRU model. Increasing t_s leads to improved performance with a lower MAE for both models. With increasing t_s, the overall training time also increases substantially. Across all values for t_s, the training time of convGRU-CNN+ is lower than the time for CNN-GRU.

Last, we present results for the needle insertion experiments shown in Fig. 6. We performed one experiment with the shielding tube and without. When using the decoupled tube, the force sensor's measurements for ground-truth annotation closely match the the values predicted by our model. Without the tube, friction forces induce a large difference between measurements and predictions.

Discussion

We present a new technique for needle tip force estimation using an OCT fiber embedded into a needle that images the deformation of an epoxy layer. OCT has been used for multiple needle-based tissue classification scenarios [17,34] which could lead to more widespread application in clinical settings. Our needle is flexible in design and easy to manufacture. This is highlighted by our results for three needles with epoxy layers of different stiffness. With increasing stiffness, the rMAE increases slightly by 30 % between needles which indicates that there is a decrease in relative performance for stiffer needles but the decrease appears to be bounded as it is similar for needles 1 and 2 and needles 2 and 3. Also, the CC remains high in a range of 0.9997 to 0.9991. This indicates that our method generalizes well for different epoxy stiffness levels. Overall, this allows for flexible adaptation of our needle to scenarios with different requirements for force sensitivity and range.

The OCT fiber within the needle produces series of A-scans that can be treated as spatio-temporal data, i.e., 1D images over time. To process this type of data, we propose a novel convGRU-CNN+ architecture. The model performs both temporal and spatial processing and outperforms the pure temporal GRU and pure spatial 1D CNN with an MAE of 1.59 ± 1.3 compared to an MAE of 3.02 ± 3.7 and 3.26 ± 3.9, respectively. Also, we compared to the spatio-

Table 2 Comparison of several architectures. Needle 1 was used for this experiment

	MAE	rMAE	CC	IT
convGRU-CNN+	**1.59 ± 1.3**	**0.0199 ± 0.0172**	**0.9997**	10.3 ± 1.5
convGRU-CNN	1.75 ± 1.5	0.0211 ± 0.0182	0.9996	9.6 ± 1.7
GRU	3.02 ± 3.7	0.0377 ± 0.0479	0.9982	**2.5 ± 0.4**
1DCNN	3.26 ± 3.9	0.0393 ± 0.0482	0.9980	6.9 ± 1.3
CNN-GRU	2.01 ± 3.2	0.0247 ± 0.0411	0.9989	9.5 ± 1.4
CNN-convGRU	2.03 ± 3.3	0.0242 ± 0.0426	0.9990	11.1 ± 1.7
2DCNN	2.11 ± 3.5	0.0255 ± 0.0438	0.9987	8.6 ± 1.5
GRU-CNN	11.79 ± 8.6	0.1253 ± 0.1001	0.9948	10.1 ± 1.4
MIP-GPM	45.38 ± 38.7	0.4820 ± 0.4114	0.7767	14.8 ± 2.2

The best values are marked in bold

Fig. 4 Boxplots of the absolute errors for the top-performing spatio-temporal deep learning models. The red line marks the median, the boxes' bottom and top line mark the 25th and 75th percentiles, respectively. Red marks above the whiskers represent outliers. Notches around the median mark comparison intervals where non-overlapping intervals between boxplots indicate different medians at 5% significance level. With respect to the frequency of outliers, consider that the errors are likely not normally distributed
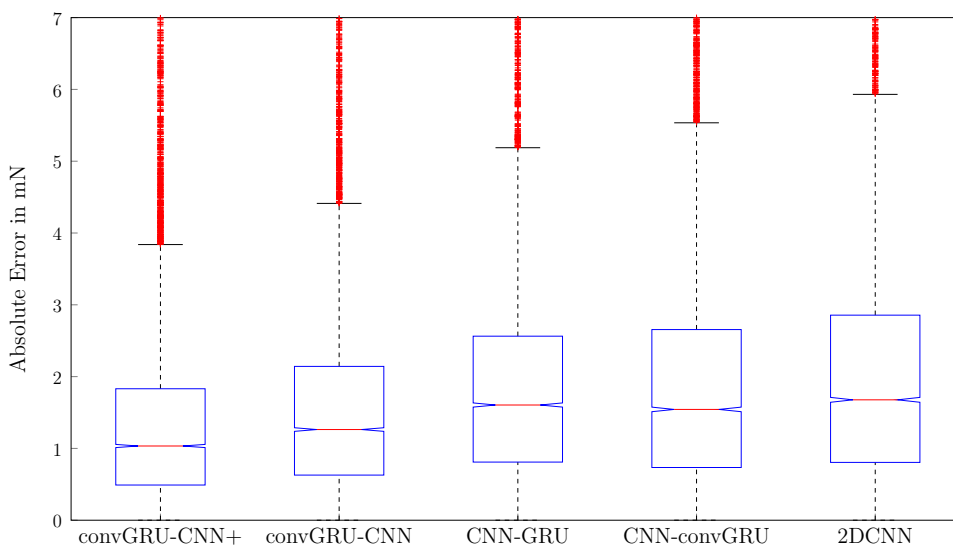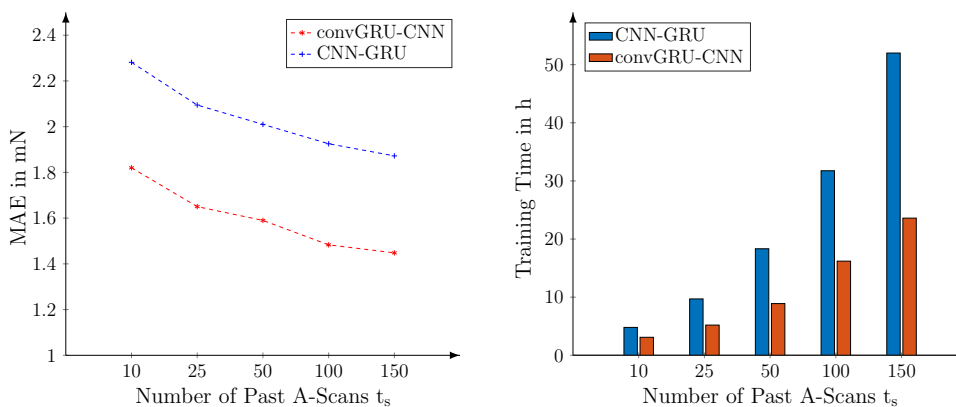

Fig. 5 Comparison between convGRU-CNN+ and CNN-GRU for different numbers of timesteps t_s. The calibration data for needle 1 was used for this experiment

temporal models CNN-GRU, CNN-convGRU and 2DCNN which are variants adopted from the natural image domain [25,26,35]. The three models are closer in terms of performance but overall, convGRU-CNN+ performs best. Notably, the differences in the median of the errors are significant which is also highlighted by the boxplots showing the test set error distribution in Fig. 4.

The key difference between all spatio-temporal deep learning models is that convGRU-CNN(+) and GRU-CNN first perform temporal processing, then spatial processing, CNN-GRU and CNN-convGRU first performs spatial, then temporal processing and 2DCNN performs concurrent processing. Overall, our proposed model significantly outperforms all other variants. The lower performance of the previous spatio-temporal models CNN-GRU [25] and

Fig. 6 Predicted and measured force values are shown for an insertion with the shielding tube (left) and without (right). For the case without tube, differences between needle tip force estimation and force sensor is caused by friction. Needle 2 and convGRU-CNN+ were used for this experiment
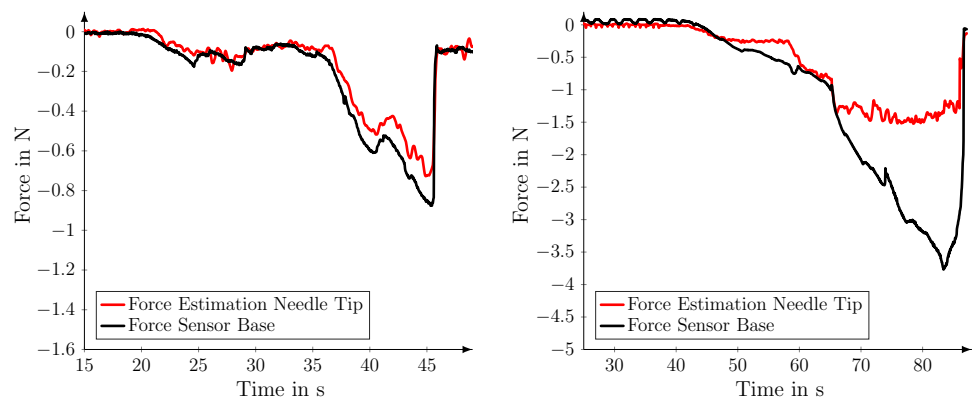
CNN-convGRU [35] indicates that temporal processing followed by spatial processing is preferable for the problem at hand. To highlight the necessity of convGRU units, we consider GRU-CNN without convolutional gates. The MAE is significantly higher which demonstrates the necessity to preserve the spatial structure during temporal processing. In addition, we show that recurrent dropout and recurrent batch normalization can improve the spatio-temporal models' performance further. For reference, MIP-GPM shows that conventional feature extraction without extensive engineering cannot match deep learning models' performance for this problem.

Furthermore, we perform a more detailed analysis of our convGRU-CNN+ model compared to the more common CNN-GRU model. The results in Fig. 5 show a decrease in the MAE when a longer history of A-Scans is considered. This highlights the value of exploiting temporal information for force estimation. However, this improvement is bought with a substantial increase in training time as the computational effort increases. For example, for convGRU-CNN+, using $t_s = 100$ instead of $t_s = 50$ previous measurements leads to a performance increase of 7 % and an increase in training time of 82 %. Training time is an important aspect to consider for application as newly designed needles will require an initial calibration, i.e., model training. If a new needle with adjusted epoxy layer for a particular force range needs to be available quickly, performance needs to be traded off against shorter training times. Overall, both models benefit similarly from the additional temporal information, however, convGRU-CNN+ trains faster. This is due to the convGRU units which have significantly fewer parameters than their GRU counterpart in CNN-GRU.

Besides performance and training time, the models' inference time is important for application and real-time feedback of forces. Overall, the high-performing spatio-temporal deep learnings models can process samples at 100 Hz which indicates real-time capability. Notably, 2DCNN is almost as fast as 1DCNN due to the fact that 2D convolutions are much more common and highly optimized in Tensorflow and CUDA. Thus, our convGRU-CNN model's inference time could improve further with software optimization as a 1D CNN is also part of the model. Also, note that inference times are hardly affected by the number of previous measurements t_s. After initial processing, the recurrent part of the models stores previous information in its cells' states and only requires one additional sample to be processed at each step.

Also, it is important how the estimated forces are transferred to the physician effectively [36]. Previous approaches used a haptic feedback device [37] or visual feedback [38] to provide the forces to the physicians. In this context, the application scenario and required force resolution are important as rough estimates might be sufficient for qualitative feedback methods. Other settings such as retinal microsurgery require highly accurate force measurements [39] where our high-performing models might be particularly beneficial. Thus, future work could examine our proposed force estimation method in the context of different application scenarios and force feedback methods.

Last, we validated our needle design in a realistic insertion experiment with a freshly resected human prostate. Our results in Fig. 6 show that the needle tip forces closely match the actual, decoupled, base measurements. While the decoupling is not perfect, we can show that our method accurately captures events such as ruptures. Without tip measurements or decoupling, large friction forces overshadow the actual tip forces. Overall, the experiments show that our method is usable for actual force estimation in soft tissue.

Conclusion

We introduce a new method to measure forces at a needle tip. Our approach uses an OCT fiber imaging the deformation of an epoxy layer to infer the force that acts on the needle tip. To map the OCT data to forces, we propose a novel convGRU-CNN+ architecture for spatio-temporal data processing. We provide an analysis of the model's properties,

and we show that it outperforms other deep learning methods. Furthermore, validation experiments in ex vivo human prostate tissue underline the method's potential for practical application. For future work, our convGRU-CNN+ architecture could be employed for other spatio-temporal learning problems.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Abolhassani N, Patel R, Moallem M (2007) Needle insertion into soft tissue: a survey. Med Eng Phys 29(4):413–431
2. Taylor RH, Menciassi A, Fichtinger G, Fiorini P, Dario P (2016) Medical robotics and computer-integrated surgery. In: Siciliano B, Khatib O (eds) Springer handbook of robotics. Springer Handbooks. Springer, Cham
3. Okamura AM, Simone C, O'leary MD (2004) Force modeling for needle insertion into soft tissue. IEEE Trans Biomed Eng 51(10):1707–1716
4. Kataoka H, Washio T, Chinzei K, Mizuhara K, Simone C, Okamura AM (2002) Measurement of the tip and friction force acting on a needle during penetration. In: International conference on medical image computing and computer-assisted intervention, pp 216–223
5. Rodrigues S, Horeman T, Sam P, Dankelman J, van den Dobbelsteen J, Jansen FW (2014) Influence of visual force feedback on tissue handling in minimally invasive surgery. British J Surg 101(13):1766–1773
6. Hatzfeld C, Wismath S, Hessinger M, Werthschützky R, Schlaefer A, Kupnik M (2017) A miniaturized sensor for needle tip force measurements. Biomed Eng 62(1):109–115
7. Beekmans S, Lembrechts T, van den Dobbelsteen J, van Gerwen D (2016) Fiber-optic Fabry–Pérot interferometers for axial force sensing on the tip of a needle. Sensors 17(1):38
8. Su H, Zervas M, Furlong C, Fischer GS (2011) A miniature mri-compatible fiber-optic force sensor utilizing Fabry–Perot interferometer. MEMS Nanotechnol 4:131–136
9. Rao YJ (1997) In-fibre bragg grating sensors. Meas Sci Technol 8(4):355
10. Roriz P, Carvalho L, Frazão O, Santos JL, Simões JA (2014) From conventional sensors to fibre optic sensors for strain and force measurements in biomechanics applications: a review. J Biomechan 47(6):1251–1261
11. Kumar S, Shrikanth V, Amrutur B, Asokan S, Bobji MS (2016) Detecting stages of needle penetration into tissues through force estimation at needle tip using fiber bragg grating sensors. J Biomed Opt 21(12):127,009
12. Kennedy KM, Chin L, McLaughlin RA, Latham B, Saunders CM, Sampson DD, Kennedy BF (2015) Quantitative micro-elastography: imaging of tissue elasticity using compression optical coherence elastography. Sci Rep 5(15):538
13. Larin KV, Sampson DD (2017) Optical coherence elastography-oct at work in tissue biomechanics. Biomed Opt Express 8(2):1172–1202
14. Otte C, Beringhoff J, Latus S, Antoni ST, Rajput O, Schlaefer A (2016) Towards force sensing based on instrument-tissue interaction. In: IEEE international conference on multisensor fusion and integration for intelligent systems, pp 180–185
15. Gessert N, Beringhoff J, Otte C, Schlaefer A (2018) Force estimation from oct volumes using 3d cnns. Int J Comput Assist Radiol Surg 13(7):1073–1082
16. Kennedy KM, Kennedy BF, McLaughlin RA, Sampson DD (2012) Needle optical coherence elastography for tissue boundary detection. Opt Lett 37(12):2310–2312
17. Otte C, Otte S, Wittig L, Hüttmann G, Kugler C, Drömann D, Zell A, Schlaefer A (2014) Investigating recurrent neural networks for OCT A-scan based tissue analysis. Methods Inf Med 53(4):245–249
18. Aviles AI, Alsaleh SM, Hahn JK, Casals A (2017) Towards retrieving force feedback in robotic-assisted surgery: a supervised neuro-recurrent-vision approach. IEEE Trans Haptics 10(3):431–443
19. Simonyan K, Zisserman A (2014) Two-stream convolutional networks for action recognition in videos. In: Advances in neural information processing systems, pp 568–576
20. Wang L, Qiao Y, Tang X (2015) Action recognition with trajectory-pooled deep-convolutional descriptors. In: CVPR, pp 4305–4314
21. Yue-Hei Ng J, Hausknecht M, Vijayanarasimhan S, Vinyals O, Monga R, Toderici G (2015) Beyond short snippets: deep networks for video classification. In: CVPR, pp 4694–4702
22. Long X, Gan C, de Melo G, Wu J, Liu X, Wen S (2018) Attention clusters: Purely attention based local feature integration for video classification. In: CVPR
23. Ji S, Xu W, Yang M, Yu K (2013) 3D convolutional neural networks for human action recognition. IEEE Trans Pattern Anal Mach Intell 35(1):221–231
24. Tran D, Bourdev L, Fergus R, Torresani L, Paluri M (2015) Learning spatiotemporal features with 3D convolutional networks. In: ICCV, pp 4489–4497
25. Donahue J, Anne Hendricks L, Guadarrama S, Rohrbach M, Venugopalan S, Saenko K, Darrell T (2015) Long-term recurrent convolutional networks for visual recognition and description. In: CVPR, pp 2625–2634
26. Xingjian S, Chen Z, Wang H, Yeung DY, Wong WK, Woo Wc (2015) Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: Advances in Neural Information Processing Systems, pp 802–810
27. Gessert N, Priegnitz T, Saathoff T, Antoni ST, Meyer D, Hamann MF, Jünemann KP, Otte C, Schlaefer A (2018) Needle tip force estimation using an oct fiber and a fused convgru-cnn architecture. In: International conference on medical image computing and computer-assisted intervention, pp 222–229
28. Cooijmans T, Ballas N, Laurent C, Gülçehre Ç, Courville A (2017) Recurrent batch normalization. In: ICLR

29. Pham V, Bluche T, Kermorvant C, Louradour J (2014) Dropout improves recurrent neural networks for handwriting recognition. In: International conference on frontiers in handwriting recognition, pp 285–290

30. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: CVPR, pp 770–778

31. Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using rnn encoder-decoder for statistical machine translation. In: EMNLP, pp 1724–1734

32. Rasmussen CE (2004) Gaussian processes in machine learning. In: Advanced lectures on machine learning, pp 63–71

33. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M et al (2016) Tensorflow: a system for large-scale machine learning. OSDI 16:265–283

34. Muller BG, de Bruin DM, Van den Bos W, Brandt MJ, Velu JF, Bus MT, Faber DJ, Heijink CDS, Zondervan PJ, de Reijke TM et al (2015) Prostate cancer diagnosis: the feasibility of needle-based optical coherence tomography. J Med Imaging 2(3):037,501

35. Ballas N, Yao L, Pal C, Courville A (2016) Delving deeper into convolutional networks for learning video representations. In: ICLR

36. Kitagawa M, Dokko D, Okamura AM, Yuh DD (2005) Effect of sensory substitution on suture-manipulation forces for robotic surgical systems. J Thorac Cardiovasc Surg 129(1):151–158

37. Meli L, Pacchierotti C, Prattichizzo D (2017) Experimental evaluation of magnified haptic feedback for robot-assisted needle insertion and palpation. Int J Med Robot Comput Assist Surg 13(4):e1809

38. Aviles-Rivero AI, Alsaleh SM, Philbeck J, Raventos SP, Younes N, Hahn JK, Casals A (2018) Sensory substitution for force feedback recovery: a perception experimental study. ACM Trans Appl Percept (TAP) 15(3):16:1–16:19

39. He X, Handa J, Gehlbach P, Taylor R, Iordachita I (2014) A submillimetric 3-DOF force sensing instrument with integrated fiber Bragg grating for retinal microsurgery. IEEE Trans Biomed Eng 61(2):522–534